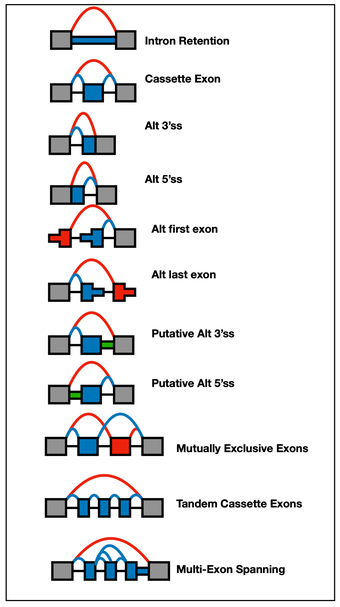
Integrative transcriptomic analysis to characterize RBPs involved in 3’ UTR alternative splicing

Evan Correa[1], Matthew Gazzara [2], Yoseph Barash [2]

1 Federated Department of Biological Sciences, New Jersey Institute of Technology
2. Department of Genetics, Perelman School of Medicine, University of Pennsylvania

Background

- Splicing is an essential processing step in converting an intron-containing pre-mRNA transcript into a mature mRNA molecule that will then be translated.
- The mature mRNA will typically consist of exonic regions that are spliced together upon removal of intronic (non-coding) regions.
- Cells can splice these pre-mRNA transcripts in various ways that can result in multiple protein isoforms, a process known as **alternative splicing**.
- We are interested in splicing events that occur near the **3’ UTR** of an mRNA molecule (Alternative last exon and putative alternative last exon).
- Dysregulation of these events are associated with a variety of cancers.
- RNA-binding proteins (RBPs) have complex regulatory functions and can bias towards specific splicing decisions.



Methods

- **ENCODE (Encyclopedia of DNA Elements) –** RNA-seq data from gene knockouts of 356 unique RBPs. Knockouts were conducted in Hepg2 and k562 cell lines.
- **MAJIQ v2** – RNA-seq was analyzed using MAJIQ v2. This program categorizes the splicing events and quantifies ΔPSI (change in Percent Spliced In) between the control and knockout. This output is further associated with specific splicing junctions that can be quantified individually.
- **Downstream Analysis –** Python scripts were written to parse through MAJIQ output to determine which RBPs were associated with significant changes in last exon events.
- **MAJIQlopedia –** Web tool was used to visualize the genes and quantify the splice junctions.

Results

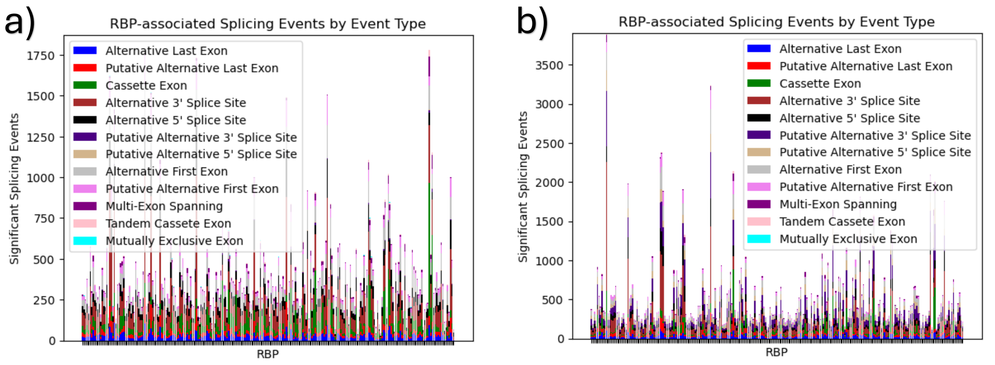


Figure 1. RBPs and their respective associations with significant splicing changes. Two stacked bar charts depicting significant splicing changes between knockout and control samples in a)Hepg2 and b)k562 cell lines. Significance was established using a 0.2 E(PSI) and 95% confidence interval. Event types were color coded and stacked to show relative abundance across different RBP knockouts.

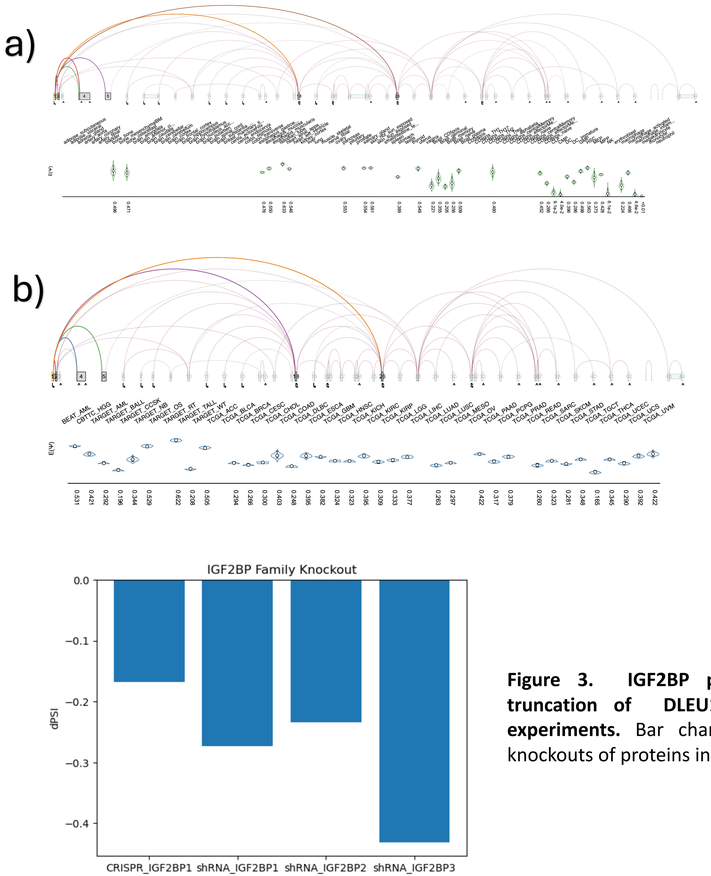


Figure 2. IGF2BP1 is associated with the truncation of DLEU1 in both normal and cancer tissue. a) MAJIQlopedia generated plot showing splice graph for DLEU1, a gene that was shown to have significant splicing changes between control and knockout of IGF2BP1 in Hepg2 cell line. Violin plot shows relative inclusion (PSI) of the truncated transcript across different tissues. The same analysis is done for b), where the splice graph is visualized for cancer samples. Violin plot shows PSI across different cancer types and solid tumors.

Figure 3. IGF2BP protein family associated with truncation of DLEU1 across 4 different knockout experiments. Bar chart showing dPSI values for 4 knockouts of proteins in the IGF2BP family.

Conclusions

- An initial analysis, solely based on the correlation of RBP knockout and splicing changes identified, we have identified **HNRNPL, HNRNPC, EXOSC5, IGF2BP1, SFPQ, NUSAP1, MAK16, TRIM56**, as candidate proteins that could have regulatory functions in the 3’ UTR.
- These results confirm that we are finding high quality targets; **SFPQ** is a known splicing factor that has direct regulatory role in alternative splicing.
- DLEU1 is an RNA-gene that is frequently deleted in patients with B-cell chronic lymphocytic leukemia.
- The knockout of IGF2BP1, as a representative example, resulted in a preference for the proximal last exon, which given its early position in the gene essentially means the RNA molecule becomes truncated.
- This preference for the proximal last exon held when analyzing cancer tissue.
- We have identified the IGF2BP protein family to have a truncating effect on the DLEU1 lncRNA.
- This effect held across 4 different knockout experiments, including a CRISPR interference knockout.

Future Directions

- Further efforts will integrate eCLIP and motif data to determine where these RBPs are binding and how they perform their regulatory roles.
- We can also perform similar analyses for different kinds of splicing events.
- We can also characterize the role of other similar protein families that regulate last exon events.

Acknowledgements

I would like to thank the IDEAL Research program for financially supporting the work I conducted this summer.

I would also like to thank Matthew Gazzara and Dr. Yoseph Barash for guiding this research project and supporting it at every stage.